



AFRL-RH-WP-JA-2010-0002

Evaluation of Eye Metrics as a Detector of Fatigue

**R. Andy McKinley
Lindsey K. McIntire
Regina Schmidt
Andrea Pinchak
John L. Caldwell**

**Biosciences and Performance Division
Vulnerability Analysis Branch**

**Daniel W. Repperger
Warfighter Interface Division
Battlespace Visualization Branch**

**Matt Kane
The Henry M. Jackson Foundation
1401 Rockville Pike, Suite 600
Rockville, MD 20852**

March 2010

Interim Report for March 2007 to March 2009

**Approved for public release;
distribution is unlimited.**

**Air Force Research Laboratory
711th Human Performance Wing
Human Effectiveness Directorate
Biosciences and Performance Division
Vulnerability Analysis Branch
Wright-Patterson AFB OH 45433**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th Air Base Wing Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RH-WP-JA-2010-0002 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//SIGNED//

Suzanne Smith, Work Unit Manager
Vulnerability Analysis Branch

//SIGNED//

Mark M. Hoffman
Biosciences and Performance Division
Human Effectiveness Directorate
711th Human Performance Wing
Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 01 03 2010		2. REPORT TYPE Interim Report		3. DATES COVERED (From - To) March 2007 – March 2009	
4. TITLE AND SUBTITLE Evaluation of Eye Metrics as a Detector of Fatigue				5a. CONTRACT NUMBER In-House	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) R. Andy McKinley*, Lindsey K. McIntire*, Regina Schmidt*, Andrea Pinchak*, John L. Caldwell*, Daniel W. Repperger**, and Matt Kane***				5d. PROJECT NUMBER 7184	
				5e. TASK NUMBER 02	
				5f. WORK UNIT NUMBER 71840223	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) *Air Force Materiel Command ***The Henry M. Jackson Foundation Air Force Research Laboratory 1401 Rockville Pike, Ste 600 711th Human Performance Wing Rockville MD 20852 Human Effectiveness Directorate Biosciences and Performance Division **Warfighter Interface Division Vulnerability Analysis Branch Battlespace Visualization Branch Wright-Patterson AFB OH 45433-7947				10. SPONSOR/MONITOR'S ACRONYM(S) 711 HPW/RHPA	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-RH-WP-JA-2010-0002	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES 88ABW/PA cleared on 28 Jan 10, 88ABW-2010-0377					
14. ABSTRACT Objectives: The purpose of this study was to evaluate oculometrics as a detector of fatigue in Air Force relevant environments using one night of sleep deprivation. Method: Ten civilian participants volunteered to participate in this study. Each was trained on three performance tasks: target identification, unmanned aerial vehicle (UAV) landing, and the psychomotor vigilance task (PVT). Experimental testing of the three tasks began after 14 hours awake, and continued every two hours until 28 hours of sleep deprivation was reached. Results: Data analyses showed statistically significant decrements in performance as the level of sleep deprivation increased, for both the PVT and the target identification task. These performance declines correlated with increases in proportion of eye closure and declines in approximate entropy of pupil position. Conclusion: The results provide evidence that eye metrics can be used to detect the onset of fatigue, potentially in advance of significant changes in operator performance, suggesting a way to predict fatigue-induced declines in performance before they manifest.					
15. SUBJECT TERMS fatigue, alertness, eye tracker, monitoring device, sleep deprivation					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			R. Andy McKinley
U	U	U	SAR	25	19b. TELEPHONE NUMBER (include area code) NA

THIS PAGE IS INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

ABSTRACT	iv
INTRODUCTION	1
METHOD	2
Participants.....	2
Apparatus	2
Stimuli.....	3
Procedure	4
RESULTS AND DISCUSSION	5
Performance Data:	5
Discussion.....	16
CONCLUSIONS AND FUTURE RESEARCH	18
REFERENCES	19
BIOGRAPHIES	21

LIST OF FIGURES

Figure 1. Average Target Acquisition Time	7
Figure 2. Average PVT Response Time	8
Figure 3. Average Number of Lapses during PVT	9
Figure 4. Target Acquisition Task. Total Eye Closure Duration Mean Percent change from Baseline.	12
Figure 5. PVT Task. Total Eye Closure Duration Mean Percent change from Baseline.	13
Figure 6. UAV Task. Total Eye Closure Duration Mean Percent Change from Baseline.	14
Figure 7. Maximum ApEn versus Session Number	16

ABSTRACT

Objectives: The purpose of this study was to evaluate oculometrics as a detector of fatigue in a Air Force relevant environments using one night of sleep deprivation. Using the eye metrics of total eye closure duration and approximate entropy, the relationship between these eye metrics and the onset of fatigue-induced performance decrements was investigated. **Background:** Perhaps, the most damaging effects to the successful outcome of operational military missions are those attributed to sleep deprivation-induced fatigue. Consequently, there is increasing interest in the development of reliable monitoring devices that can easily assess when an operator or soldier may be overly fatigued and operating at a dangerous performance level. **Method:** Ten civilian subjects volunteered to participate in this study. Each was trained on three performance tasks: target identification, unmanned aerial vehicle (UAV) landing, and the psychomotor vigilance task (PVT). Experimental testing of the three tasks began after 14 hours of sleep deprivation, and continued every two hours until 28 hours of sleep deprivation was reached. **Results:** Data analyses showed statistically significant decrements in performance as the level of sleep deprivation increased, for both the PVT and the target identification task. These performance declines correlated with increases in proportion of eye closure and declines in approximate entropy of pupil position. **Conclusion:** The results provide evidence that eye metrics can be used to detect the onset of fatigue, potentially in advance of significant changes in operator performance, suggesting a way to predict fatigue-induced declines in performance before they manifest. **Application:** Operators in both manned and unmanned vehicle environments, as well as other military or commercial operator environments, could benefit greatly from an alertness monitoring device for assessing fatigue status and task performance.

INTRODUCTION

The operational military environment presents a variety of environmental and physiological stressors such as vibration, heat, high acceleration, and fatigue that can significantly influence the effectiveness and task performance of the individual. Perhaps the most damaging effects to the successful outcome of the mission are those attributed to sleep deprivation-induced fatigue. Although considerable research has been conducted to assuage the effects of pilot fatigue, it remains a significant concern in aviation operations. NASA's Aviation Safety Reporting System (ASRS) routinely receives reports from pilots blaming fatigue, sleep loss, and sleepiness in the cockpit for operational errors such as altitude and course deviations, fuel miscalculations, landings without proper clearances, and landings on incorrect runways (Caldwell, Caldwell, and Schmidt, 2008; Rosekind et al., 1994). Aviator fatigue is associated with degradations in response accuracy and speed, the unconscious acceptance of lower standards of performance, impairments in the capacity to integrate information, and narrowing of attention that can lead to forgetting or ignoring important aspects of flight tasks (Perry, 1974). As sleepiness levels increase, performance becomes less consistent and vigilance deteriorates (Dinges, 1990).

There is a need to accurately determine when pilots become fatigued to dangerous levels, but progress in the area of "fitness for duty" testing and "real-time monitoring" of operator performance has been slow (Institute of Medicine, 2004). Oculometric-based techniques appear as a promising measure for fatigue. Stern (1999) has shown that oculomotor measures are useful for detecting lapses in attention, and several other investigators (Dinges et al., 1998 and Russo et al., 1999) have established the sensitivity of the oculomotor control system to detect fatigue, boredom, and lapses in attention. Russo et al. (2003) reported that saccadic velocity is particularly sensitive to an increase in sleepiness in response to prolonged periods of partial sleep deprivation, and Russo et al. (1999) found that both decreases in saccadic velocity and increases in pupil constriction latency correlated with an increase in the rate of crashes under simulated driving conditions during periods of sleep deprivation. Long duration eye closures during eye blinks also have been found to provide a good indication of reduced alertness (Stern, 1980), and others have demonstrated that an integrated measure of the degree of eye closure over a specific interval of time offers information about the level of operator sleepiness (Wierwille, 1999). Specifically, Dinges et al. (1998) found a high degree of coherence between slow eye closures (measured via PERCLOS [Percent Eye Closure]) and performance lapses on a test of sustained attention, and Mallis et al. (2000) found that PERCLOS feedback from the device improved alertness and driving performance, especially when drivers were drowsy at night. In part, this is why the Federal Highway Administration and the National Highway Traffic Safety Administration consider PERCLOS to be among the most promising known real-time measures of alertness for in-vehicle drowsiness detection (Dinges and Grace, 1998).

METHOD

The goal of this experiment was to determine if eye metrics could detect fatigue before degradations in performance.

Participants

An Institutional Review Board (IRB) approved experimental protocol was used in this study. A total of 10 volunteer subjects (9 male, 1 female) were admitted to the study after providing informed consent to participate. They ranged in age from 18 to 42 years. Female subjects that may have been pregnant (determined by a urine pregnancy test) were not permitted to participate.

Apparatus

EC6 Eye-Tracker: The study required each subject to wear the EyeCom (Reno, NV) eye-tracker (EC6) during testing. The device consists of two infrared (IR)-sensitive cameras and a linear array of IR-illuminating light emitting diodes (LEDs) mounted on a set of eyeglass frames. The cameras are angled upward toward the eyes and extract real-time pupil diameter, eye-lid movement, and eye-ball movement. The software records a variety of measurements including pupil position, eye closure, size, etc. which can then be used to calculate eye-blink duration (EBD), eye-blink frequency (EBF), eye-blink velocity (EBV), percentage of time the eyes are closed (PERCLOS), saccadic eye movement velocity, and pupil response latency to light flashes. Only the left eye was monitored for this experiment as left and right eye comparisons were not necessary for the planned analyses.

Actigraph Monitor: Two days prior to data collection, each subject donned a wrist activity monitor (WAM; Ambulatory Monitoring, Inc.). The WAM is a non-invasive small electronic device that can be worn on the wrist like a wristwatch. It records limb and body movements to determine when a subject is active and when they are asleep. It was used to ensure the subjects received at least 7 hours of sleep per night for the two nights prior to data collection.

Simulation Facilities: Two simulation facilities were used in this study. The first was the Predator RQ-1 Ground Control Station (GCS). This simulation apparatus utilizes two 19-inch cathode ray tube (CRT) monitors to display a map view and a nose-camera view for the simulation. The positions of the monitors were the same as that of the operational Predator GCS (top monitor at 15° downward angle 44.5-inches above the table, bottom monitor perpendicular with the table at eye height). An overhead map view was displayed on the top monitor, and a camera view (from the nose of the simulated aircraft) was displayed on the bottom monitor, as is the case for the operational GCS. The seat utilized was a traditional office chair with an adjustable seat height (similar to the chair used by Predator pilots). A replicated Predator GCS flight stick and throttle developed by High Rev Simulators (Lancaster, CA) served as the flight control interface. The landing task included one pre-determined flight path the subject was required to follow each trial. The flight software was a modified version of X-Plane Flight Simulator, version 7.0 (Laminar Research, Columbia, SC), run on a Gateway Pentium IV desktop PC (Gateway Inc., Irvine, CA). An experienced pilot consultant ensured that the flight dynamics and displays were similar to those employed by the Predator Unmanned Aerial Vehicle UAV. X-plane provided output data that were recorded by a custom

program written in Microsoft Visual C++ (Microsoft Corp, Seattle, WA). This output data included glideslope path, altitude, airspeed, descent rate, heading, and position. Additional software was coded in OpenGL (Silicon Graphics, Inc, Mountain View, CA) and interfaced with the X-plane program to display the map and heads-up display (HUD) symbology overlay.

The second simulator was the Synthesized Immersion Research Environment (SIRE), which is a state-of-the-art virtual environment research laboratory. This facility contains a 40-foot diameter dome that serves as a high-resolution, large field-of-view (70 degrees vertical by 150 degrees horizontal) interactive visual display. A simulated UH-60 helicopter cockpit developed by Protobox, Inc. (Dayton, OH) resides in the center of the dome display. The cockpit includes both the left and right seats, instrument cluster, cyclic, collective, and pedal controls, and a simulated UH-60 surrounding structure. This station also included an electro-hydraulic control loader system to raise the simulated UH-60 cockpit to the desired height. Subjects participating in the study were required to perform a target identification task within a simulated flight environment. Although the simulation is capable of being flown manually by a human operator, all flights used in this experiment were computer controlled due to the fact the subjects were simply required to visually search for targets.

Psychomotor Vigilance Task (PVT): The PVT-192 psychomotor vigilance task (Ambulatory Monitoring, Inc.; Ardsley, NY) is a handheld, computerized test presentation and data capture system that records simple reaction times on a visual detection task, known to be sensitive to sleep loss (Dinges et al., 1997). The visual stimulus is presented on a small liquid crystal display (LCD) and subject responses are secured using two separate buttons. The PVT requires sustained attention and discrete motor responses. The 8" x 4.5" x 2.4" portable, battery-operated device visually displays numbers counted up by milliseconds in a window.

Grass Telefactor Electroencephalograph: Each subject was instrumented with electroencephalograph (EEG) sensors to monitor brain activity during testing. The sensors were attached to the scalp with medical grade adhesive tape. The Federal Drug Administration (FDA)-approved Grass Telefactor (Astro-Med, Inc.; West Warwick, RI) EEG system includes an amplifier and a computer to record and analyze the data. The amplifier can accept up to 32 sensors and samples data up to 400 hertz (Hz). Mindline channels Fp₁, C₃, C₄, F_z, C_z, P_z, and O_z from the standard 10-20 sensor placements system were selected for the montage used in this experiment and were referenced to electrode sites A₁ and A₂ (located behind the ears) during recording. The time constant for the EEG channels was 0.3 second, and the high-pass filter was set to 70 Hz.

Stimuli

Subjects were required to complete three separate performance tasks for this study: target identification, UAV landing, and PVT. The target identification task was performed in the right seat of the UH-60 helicopter simulator within the SIRE facility. The simulated aircraft was computer controlled and flew straight and level at 5000 ft. Weather conditions were set to "clear" and time of day was 12:00 pm to optimize visual conditions and allow subjects to readily see each target aircraft.

During each trial, three target aircraft were simultaneously presented at a simulated distance of 200 ft from the subject's aircraft. The aircraft types for the three targets were randomized with the

condition that at least two of the aircraft must be different. A total of fifty trials were administered in each of seven sessions. Each target's elevation and azimuth were randomly generated within the limitations of the SIRE screen field of view. Elevation was limited to $\pm 20^\circ$ whereas azimuth was limited to $\pm 50^\circ$. Three types of target aircraft were displayed including enemy (Su-37), friendly (F-22), and unknown (F-16). These aircraft were chosen due to their similarity in size ensuring equal salience on the screen. Subjects were required to visually acquire each target by fixating a laser point beam (laser pointer was attached to the top of their helmet) in the center of the target (± 2 degrees of visual arc) for 4 seconds. Head tracking was accomplished using a Polhamus head tracker (Colchester, VT). Subjects were instructed to acquire the targets in the order of all enemies, all unknown, and then all friendlies.

The simulated UAV landing task was generated in the X-plane software (Laminar Research, Columbia, SC) environment. The subjects flew a simulated MQ-1 Predator that was positioned at 800 ft of altitude near the main runway located at the Indian Springs Airport (Creech Air Force Base) in Nevada. Three waypoints were situated in the terrain database. The first waypoint was at 800 ft of altitude, the second was located at 600 ft, and the third was set to 360 ft. Two of these waypoints were indicated on the map with red dots and the third was positioned 0.9 mile out from the runway (in line with the runway). The standard MQ-1 Predator symbology was overlaid on the nose camera monitor. The task required each subject to fly through each waypoint at the specified altitude and land successfully on the runway. Successful landings were quantified by using three performance parameters. Subjects were instructed to achieve a glideslope of less than 20 feet, a touchdown vertical velocity less than 220 feet per second, and a touchdown airspeed of under 70 knots. To increase the difficulty, visibility was set to 0.8 nautical miles (NM) and the time of day was set to midnight.

The PVT task consisted digital number presented on a small LCD screen that started at zero and counted up sequentially by milliseconds (up to 60,000 ms, or 1 min) until the response button was pressed. When the subjects detected the start of the task, they pressed a microswitch that recorded the reaction time to the stimulus and cleared the screen for the next trial. If the subjects failed to detect task, a time-out was recorded. The interstimulus interval varied randomly from 2 to 12 seconds. The data were stored on computer and processed by custom software for subsequent analysis.

Procedure

Subjects were provided training on all three performance tasks over a period of at least three separate days. Training was completed once their task performance remained within 10% of their previous training session. Extra training was administered when necessary. Data collection was held the next day after they were trained. Refresher training was only given if they missed the data collection appointment. Two days prior to experimental trials, participants were given an activity wrist monitor and instructed that their daily schedules should include a minimum of seven hours sleep per night.

On the testing day, subjects were required to awaken at 0700 hours (7:00 AM) and perform their daily activities as normal. They were instructed to not consume any caffeine or central nervous system (CNS) altering medications/substances on the experimental test day. Each participant arrived at the test facility at 2000 hours (8:00 PM) and their activity data was analyzed to verify that proper sleep

cycles were maintained. Upon verification, the subject was instrumented with the EC6 monitoring equipment and EEG sensors to record brain activity. At 2100 hours (9:00 PM), subjects began the first of 7 experimental test sessions. They completed one entire battery of tasks which included one session of the target acquisition task (30 minute duration), one session of the UAV landing task (30 minute duration), and one session of the PVT (10 minute duration). Additionally, resting EEG data were collected with the subjects' eyes open (2 minutes) and closed (2 minutes) both before and after each task. This provided a baseline for the task they were about to begin. They were then provided with a 50-minute break of free time during which they could participate in any activity except for sleeping. This procedure was repeated once every 2 hours for 7 repetitions. The experimental test day concluded at 1100 hrs (11:00 AM) the following morning. Testing started at 14 hours of total sleep deprivation, and continued until the last session, which started at 26 hours of total sleep deprivation. After the last session, the subject was brought to his/her quarters by a well-rested member of the staff.

RESULTS AND DISCUSSION

Performance Data:

Target Acquisition Task: The target acquisition data files provided an acquisition time for each of the targets in each trial. These acquisition times were then averaged across targets and trials for each of the seven sessions within each subject. Next, the value for Session 1 was denoted as the baseline value where the subject was rested and operating normally. The remaining averages were normalized across subjects by calculating the acquisition times as a percentage change from baseline. A one-way ANOVA was performed using these normalized values. The results showed a significant main effect of session on reaction time ($F(6,63)=3.707, p=.003$). A Bonferroni post-hoc test ($\alpha=.05$) was used to examine differences between the seven sessions. The results found a significant increase in response time ($p=.023$) from baseline (Session 1) to Session 6 (Figure 1). Increases in response time indicate an increased latency and therefore a degradation in performance. No significant results were found for the other sessions. The average percent change from baseline performance was 28.131%.

Simulated UAV Landing Task: The UAV landing task assessed landing performance using glideslope root mean square error (RMSE), vertical velocity at touchdown, and airspeed at touchdown. The optimal glideslope of 4° was calculated from the final waypoint to the touchdown point on the runway. RMSE was calculated based on deviations from this optimal line. A one-way ANOVA was calculated for each variable. There was not a significant effect of session on any of the three performance metrics.

Psychomotor Vigilance Task (PVT): Performance on the PVT was assessed via three dependent variables: reaction time, lapses (reaction times greater than 5000 ms), and number of false starts. A one-way ANOVA was performed for each of these performance metrics. There was a significant main effect for session on the reaction times ($F(6, 56)=3.652, p=.004$). Specifically, the Bonferroni pairwise comparison ($\alpha=.05$) between Session 1 and Session 6 exhibited a significant difference ($p=.018$). The mean reaction time for Session 1 was 239.8 msec (se = 14.2 msec) whereas the mean for Session 6 was 310.7 msec (se = 14.2 msec) (Figure 2).

There was a significant difference in the number of lapses across sessions ($F(6,56)=3.672$, $p=.004$). Session 6 had a significantly higher number of lapses than both Session 1 ($p=.015$) and Session 3 ($p=.043$) (Figure 3) according to the Bonferroni pairwise comparisons ($\alpha=.05$). No significant differences in the number of false starts were detected across the sessions.

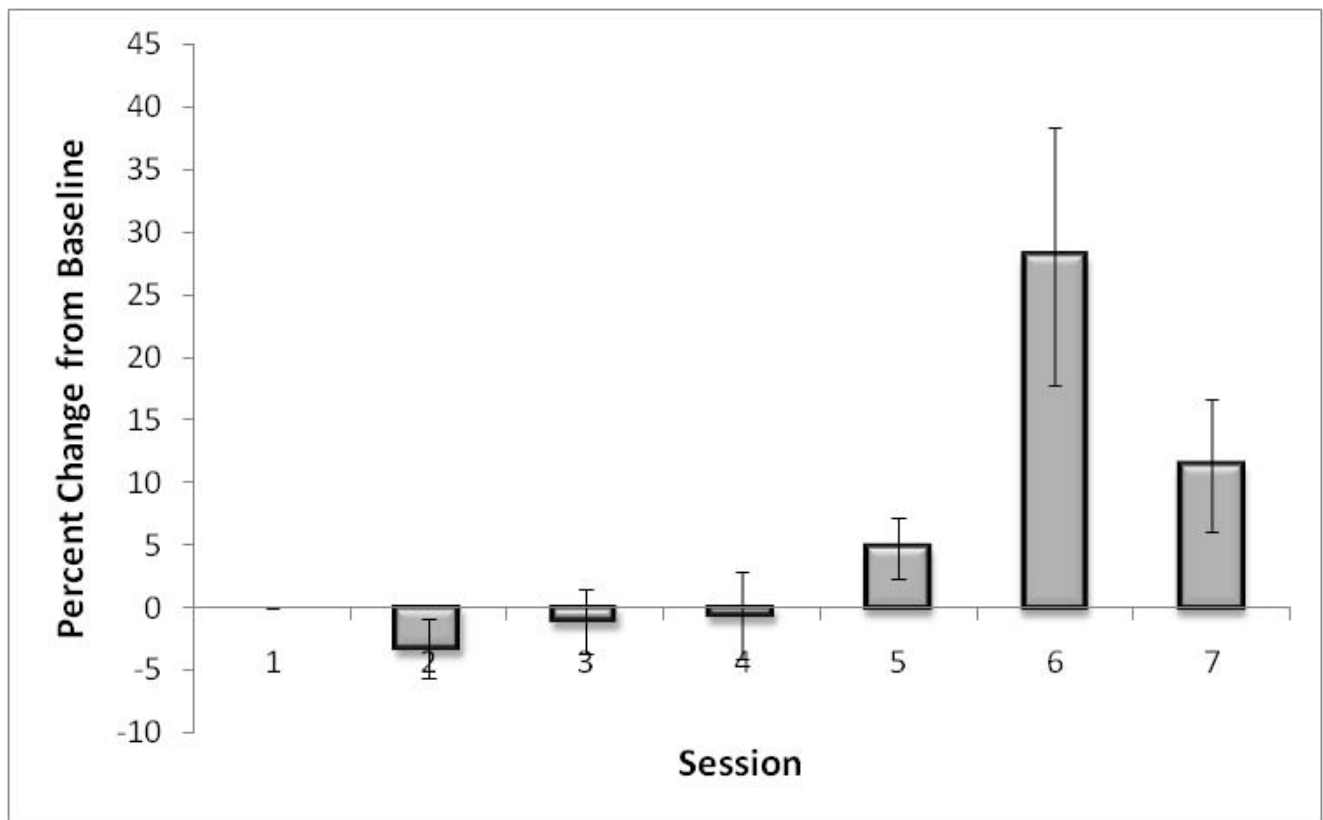


Figure 1. Average Target Acquisition Time

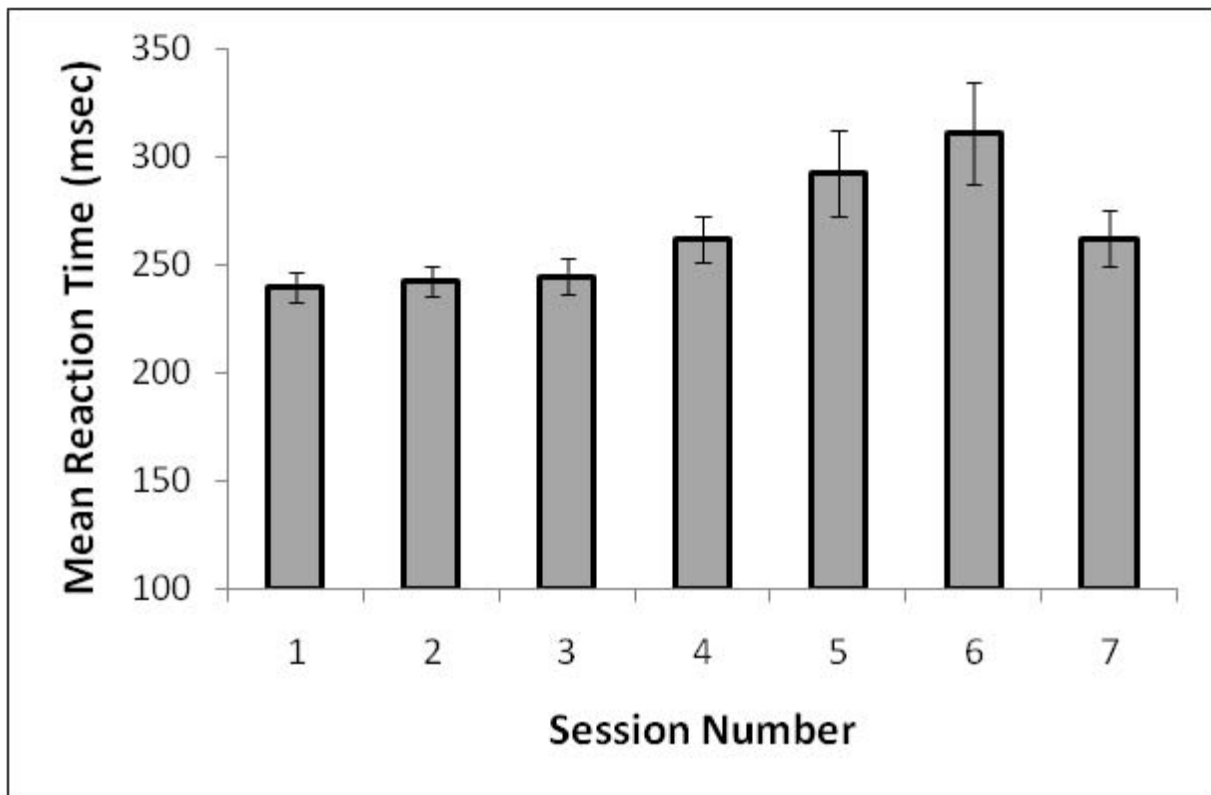


Figure 2. Average PVT Response Time

Figure 3

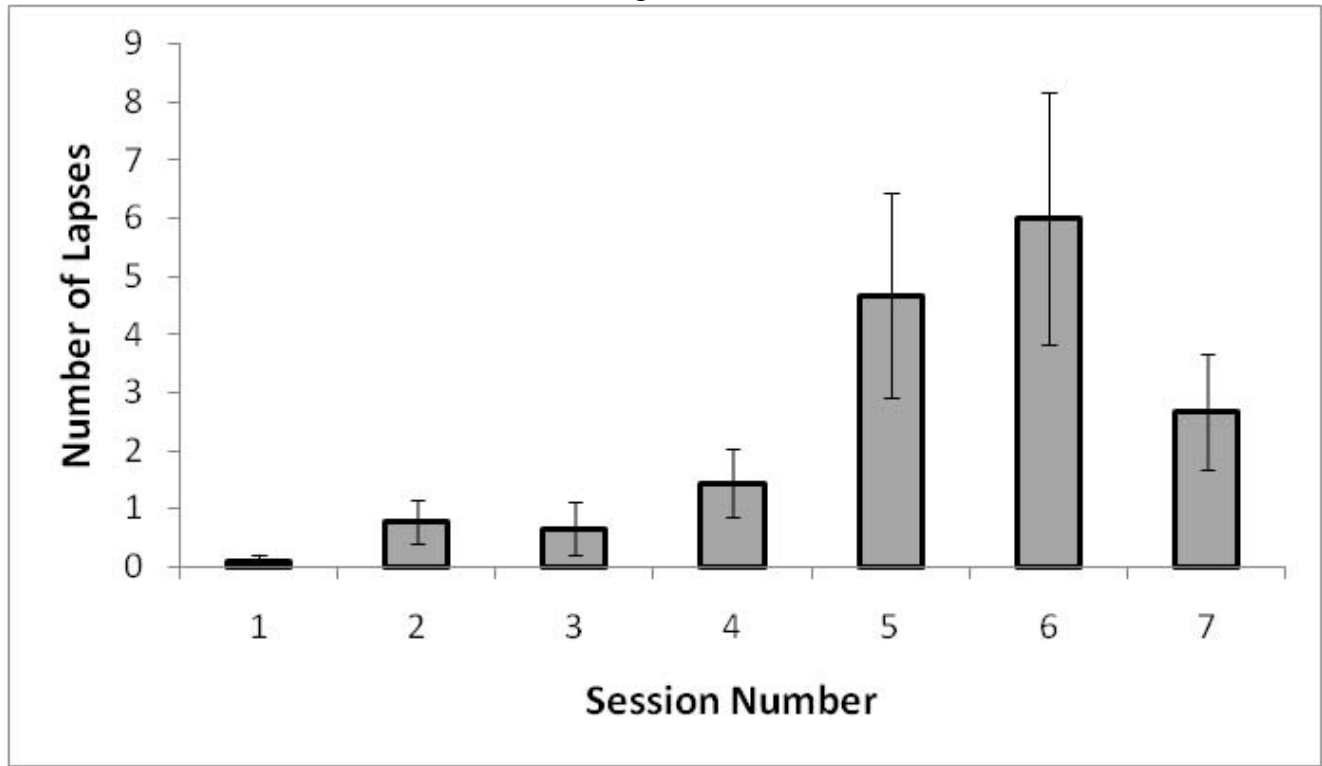


Figure 3. Average Number of Lapses during PVT

EEG: EEG data were classified into the four standard activity bands of delta (1.5- 3 Hz), theta (3. 0-8.0 Hz), alpha (8.0-13.0 Hz), and beta (13.0-20.0 Hz) by calculating the power spectrum (using a Hamming window) on a minimum of three 3-second epochs within each EEG segment of interest. This procedure was used for both the EEG collected during the PVT as well as for the resting EEGs collected prior to and immediately following execution of each task. Bands were analyzed separately in a repeated measures ANOVA, with session as the repeated variable.

For the target acquisition task, there was a significant interaction between session, eyes open rest condition, and the eyes closed rest condition for the alpha frequency band power at site C_z (top, center of head) ($F(1,6) = 2.70, p < .05$). C_z alpha power decreased during the eyes closed condition as the level of sleep deprivation increased. There was a significant main effect of session on the pre and post PVT task mean alpha power at site C_z . The pre-PVT alpha power was 4.273 (se = 1.874) while the post task mean was 3.700 (se = 1.729). In addition, a significant main effect of session on mean alpha power at site P_z , located of the medial parietal cortex, was found for the UAV task ($F(6) = 3.80, p < .05$). Alpha power declined as the level of sleep deprivation increased.

Shifts in delta power were measured for the target acquisition task and PVT. There was a significant main effect of session on the pre and post target acquisition task mean delta power at site C_4 . The pre-task mean was 0.867 (se = 0.266) while the post task mean was 1.028 (se = 0.224). There was a significant main effect of session on the pre and post PVT mean delta power at site C_z . The pre-task mean was 1.346 (se = 0.238) and the post-task mean was 1.747 (se = 0.248). There was a

significant main effect of session on the mean delta power at site F_z during the PVT task $F(1, 6) = 5.15, p < .01$).

The EC6 eye-tracker developed by EyeCom, Inc., recorded eye pupil position, eye state (open or closed), and pupil size. Eye closure data were extracted from the final data files and used to determine the total amount of time the eyes were closed during each task as a proportion of the total elapsed time to complete the task. The proportion function can be found in Equation 1 where PEC denotes the total proportion of time the eyes were closed, t_c is the total duration of eye closure, and t_{total} refers to the total amount of time to complete the task:

Equation 1. Proportion of Session Time with Eyes Closed

$$PEC = \frac{t_c}{t_{total}}$$

To normalize the data across subjects, the time duration data were calculated as a percentage change from the baseline value (Session 1). The equation is presented below:

Equation 2. Percent Change from Baseline

$$\% \text{ Change} = \frac{t_i - \bar{t}}{\bar{t}} \times 100$$

where t_i is the total eye closure time of the i^{th} session and \bar{t} is total eye closure time for the baseline session. A 1-way ANOVA was conducted for each of the three performance tasks using the mean percent change in eye closure time. No significant main effects or interactions were found ($p > 0.05$). Using the subject means, standard 2-tailed t-tests were then calculated between the baseline condition (Session 1) and the remaining six sessions. This procedure was replicated for the PVT and UAV landing tasks. The results from data gathered during the target acquisition task showed mean differences from baseline for Sessions 3-7, which were all significantly different from 0 ($p = 0.329, 0.104, 0.011, 0.013, 0.046$ for sessions 3, 4, 5, 6, and 7, respectively). These values are plotted in Figure 4. The total eye closure time increased with the level of sleep deprivation (over 3.5 times the baseline value by session 7). Figure 5 illustrates the eye closure duration percent change from baseline for the PVT task. Mean differences from baseline for sessions 3, 4, 6, and 7 were found to be significantly greater than 0 ($p = 0.038, 0.022, 0.043, \text{ and } 0.036$, respectively). Additionally, session 5 approached significance ($p = 0.063$) and may have been significant with a greater n-size. Figure 6 presents the eye closure duration percent change from baseline values collected during the UAV landing task. Although the trend appears to increase with higher levels of sleep deprivation, the mean differences from baseline were not significantly different than 0.

Approximate Entropy: EC6 eye tracker data collected during the target acquisition task were analyzed using a technique known as approximate entropy. Approximate entropy (ApEn) is based on the simple principle that if a time series signal can be compared with itself (heart beat data, for example) and the amount of the disorder in a comparison or change (entropy) between relative time shifts of these data is increasing, this is probably an indication of some change in the physiological

state. This differs from traditional correlation measures since they are not based on information theory concepts and may require fixed implicit models. The use of ApEn is “model independent” and only depends on the real time data series. There are at least four reasons why approximate entropy provides new information about system complexity not normally derived from typical statistical measures (first and second order moments):

- (1) If data are noisy, the approximate entropy measure can be compared to the noise level in the data to determine what quality of true information may be present in the data.
- (2) If the data have an artifact, this does not impact the approximate entropy measure as much as it would affect typical first and second order statistical moments from the data.
- (3) Approximate entropy can be designed to work for small data samples ($n < 50$ points) and can be applied in real time, on line. Thus, changes in the state of a physical process may be quickly determined.
- (4) For pure stochastic processes, approximate entropy will become practically infinite.

Thus, the quality of the information in a signal can then be quantitatively evaluated by comparing the entropy level of the measured signal with its underlying (non random) signal component. The statistical analysis of the pupil position ApEn data consisted of a one-way repeated measures ANOVA. The independent variable was session number. The dependent measure was the ApEn values and the data were considered across subjects.

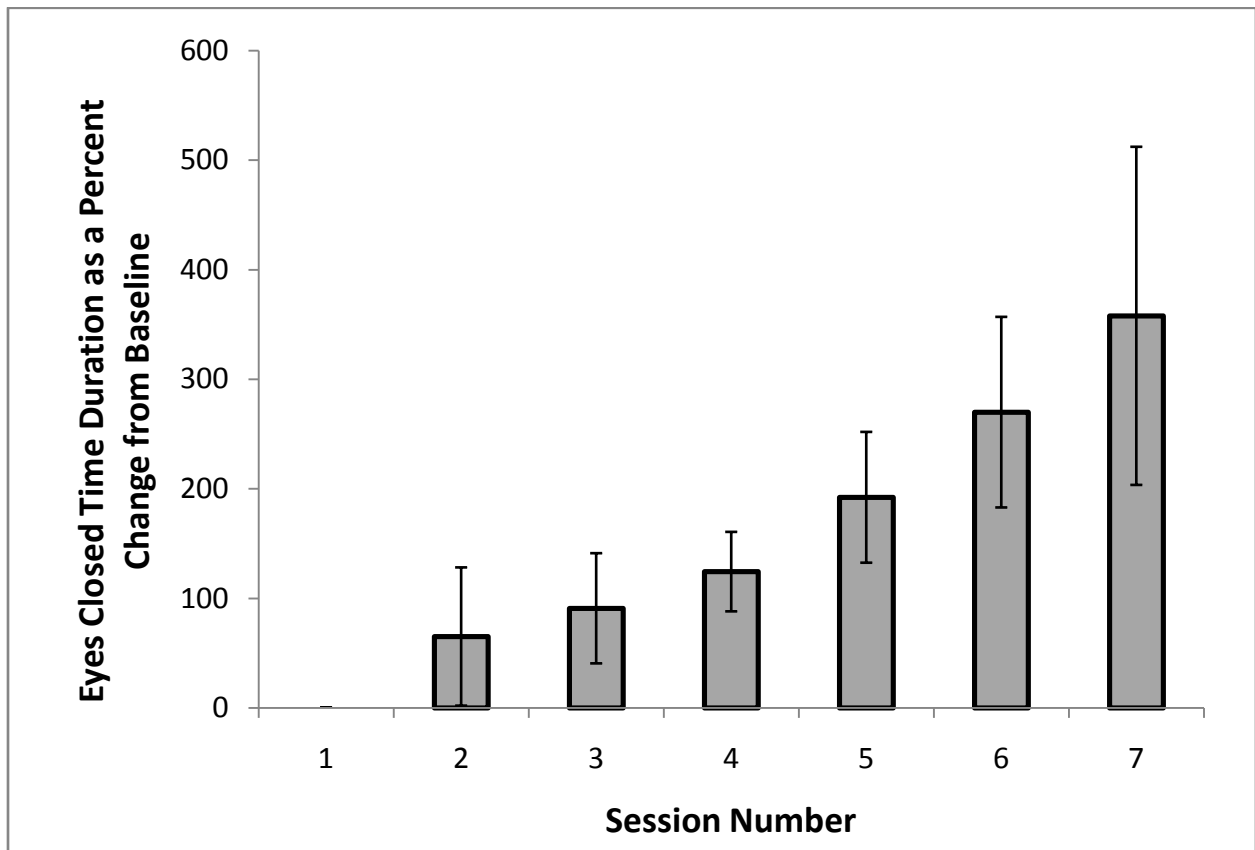


Figure 4. Target Acquisition Task. Totay Eye Closure Duration Mean Percent change from Baseline.

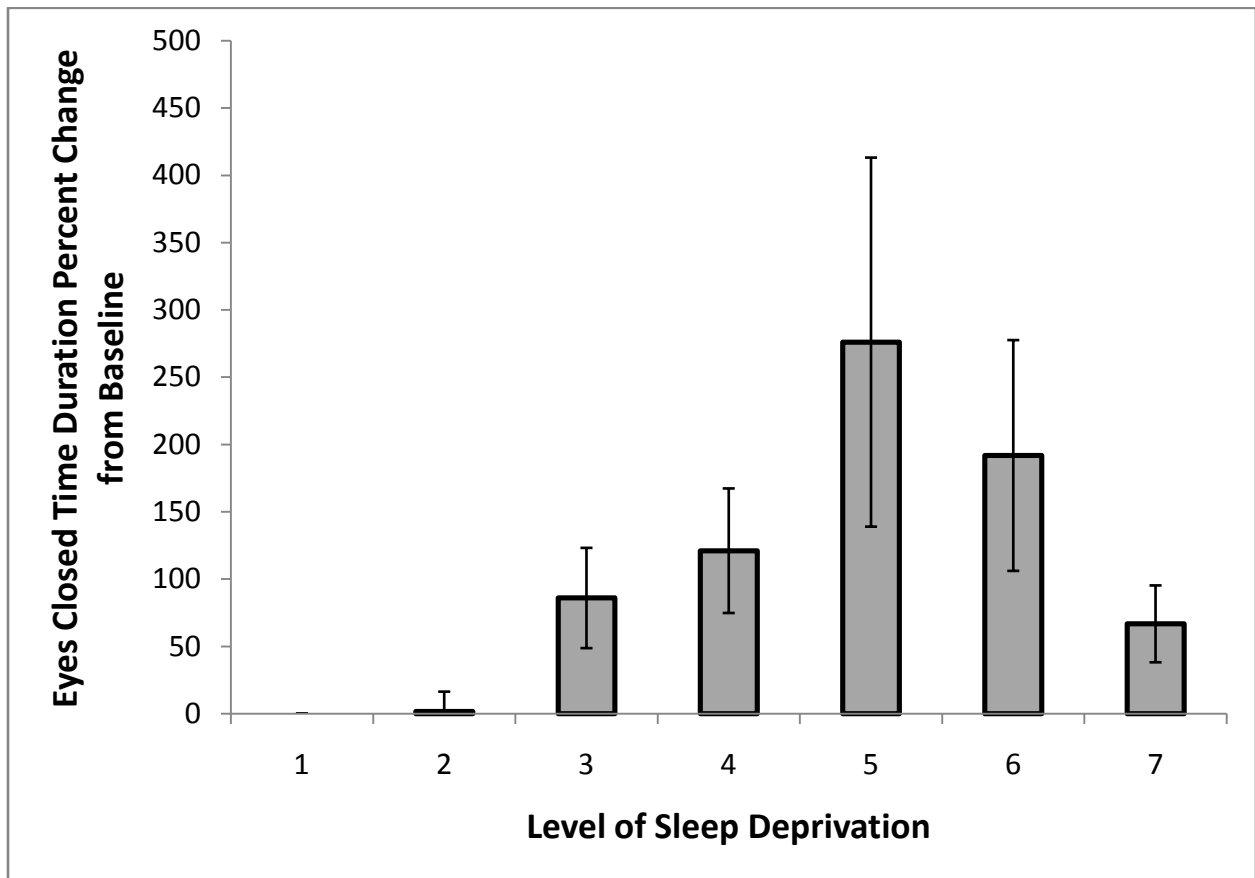


Figure 5. PVT Task. Total Eye Closure Duration Mean Percent change from Baseline.

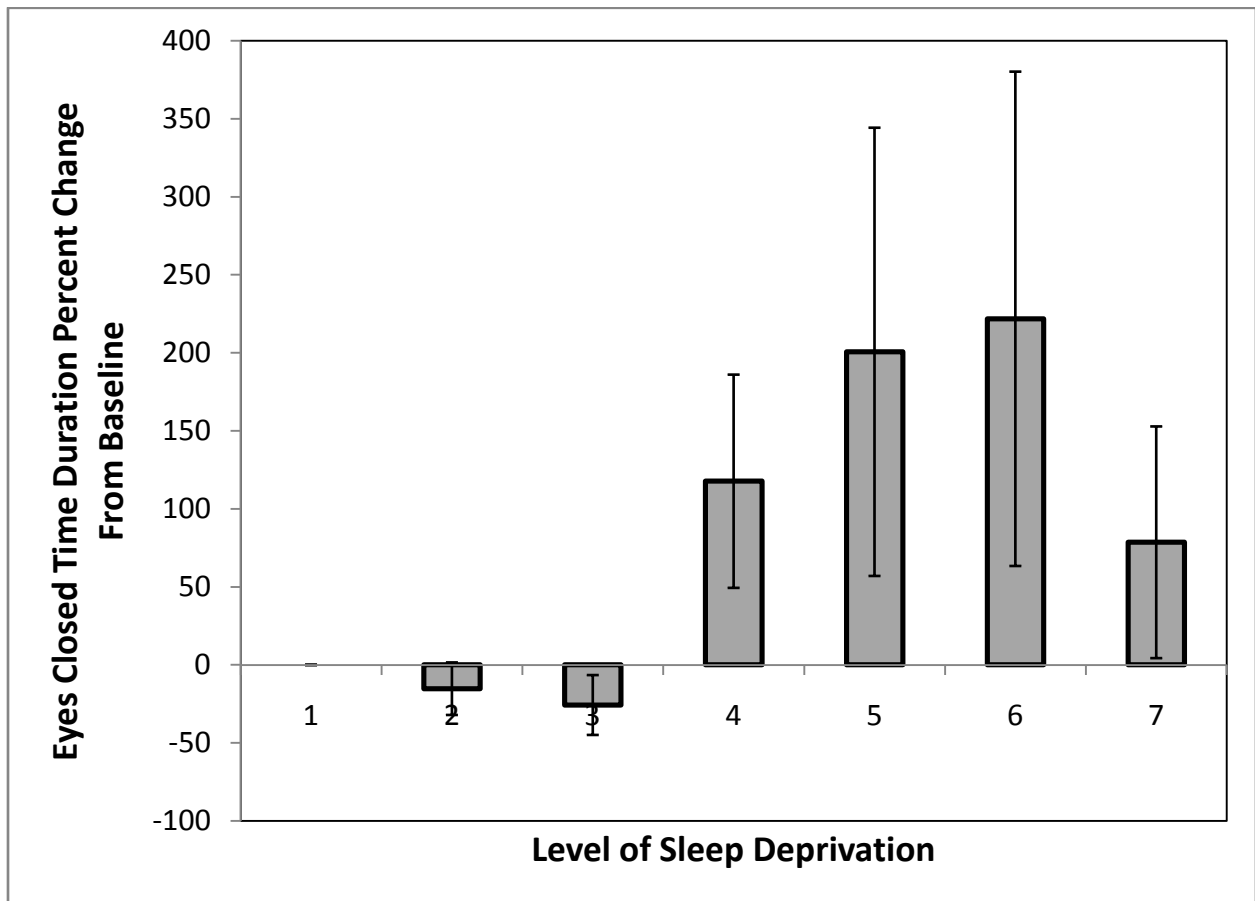


Figure 6. UAV Task. Total Eye Closure Duration Mean Percent Change from Baseline.

The data used in the ApEn analysis were the root mean square eye position about a center point. To perform the ApEn analysis, a block of run length (m) and a tolerance window (r) must be specified to compute ApEn. The parameter r is typically 20% of the standard deviation in the data (Pincus and Kalman, 2004). The $m=1$ case was applied to the fatigue data, which simply specifies the determination of the irregularity of the data between the series $s(t)$ and $s(t+1)$ and with $s(t-1)$. Since the present code was developed for 10 consecutive data points and the data from the fatigue study consisted of over 20,000 data points (33 Hz sampling rate for over 600 seconds), a Monte Carlo approach was synthesized. Each of the runs of 20,000 points was sampled 200 times uniformly and the ApEn measure was calculated yielding 200 samples of ApEn for a single run. The mean of these sampled values as well as the maximum ApEn were determined from these 200 samples. Using the sessions 2, 5 and 6, Table 1 portrays the maximum ApEn from the 200 samples within each run for all ten subjects.

Figure 7 portrays the distribution of maximum ApEn values across sessions 2, 5 and 6. It is clear that as the session number increases (going into fatigue), there is lower mean ApEn indicating less irregularity in the eye movement data. A Tukey-Kramer test was performed at an alpha of 0.05 indicating that Sessions 5 and 6 were significantly different from Session 2 but Sessions 5 and 6 were not statistically different each other. The one-way ANOVA results show a significant effect for session ($F(2,27)=10.3641, p=.0005$).

Table 1. Maximum ApEn
by subject number and session number

Subject Number	Session 2	Session 5	Session 6
Subject 1	0.965663	0.802347	0.321888
Subject 2	0.643775	0.643775	0.0602
Subject 3	0.965663	0.643775	0.643775
Subject 4	0.643775	0.321888	0.321888
Subject 5	0.940977	0.643775	0.321888
Subject 6	0.965663	0.321888	0.321888
Subject 7	0.940977	0.643775	0.802347
Subject 8	0.643775	0.321888	0.321888
Subject 9	0.802347	0.643775	0.643775
Subject 10	0.643775	0.321888	0.643775

Figure 7

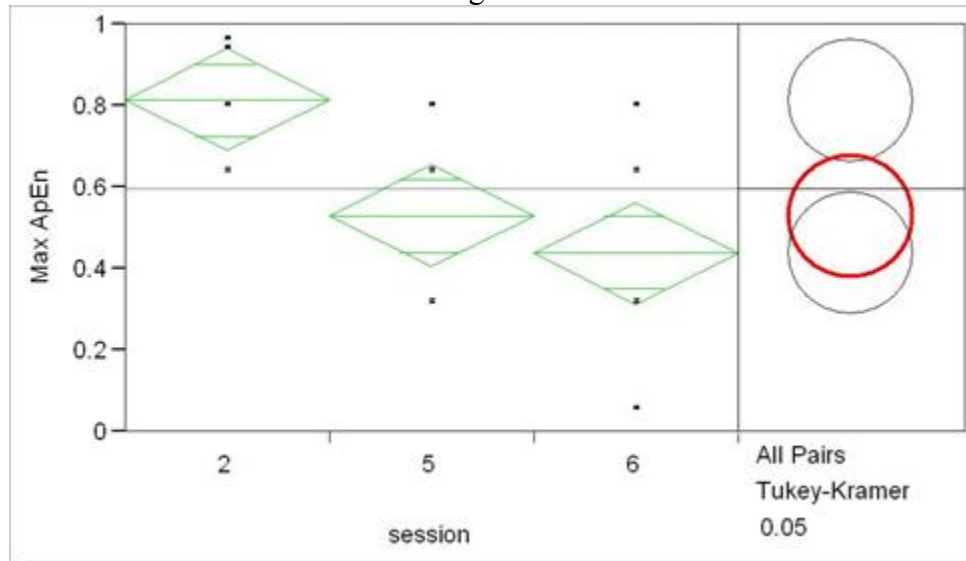


Figure 7. Maximum ApEn versus Session Number

Discussion

The main objective of this experiment was to evaluate the ability of eye metrics to predict fatigue related performance declines in relevant air operation environments, it is important to first verify that such environmental stressors were achieved. Perhaps the most well established method of objectively determining the onset of fatigue is through EEG analysis. Typically, the EEG of rested, alert individuals is comprised mainly of beta waves (13-30 Hz), although alpha activity (8-12 Hz) will dominate when the individual is calm or has eyes closed. Conversely, theta activity (4-8 Hz) becomes dominant when the subject is entering early stages of sleep and delta activity (<4 Hz) is prevalent during deep sleep (stages 3 and 4). The EEG analyses illustrate that alpha activity decreased while the delta activity increased with growing levels of sleep deprivation. It can be objectively concluded that participants experienced fatigue, particularly in Sessions 5 and 6. Perhaps due to the body's circadian rhythms and the subject's anticipation of the conclusion of the experimental data collection day (and therefore impending rest cycle), the EEG recovered somewhat for Session 7 (0900-1010) (Schmidt and Collette, 2007).

Objective performance from the fatigue study during both the target acquisition and PVT tasks followed the EEG frequency content shifts with significant declines in Session 6 and slight recovery for Session 7. The magnitude of the performance decrement also appears to depend on the type of task performed. Although significant declines in objective performance measures were found in both the target acquisition task and the PVT, the UAV landing task performance appeared relatively devoid of any fatigue consequences. It is likely that this was due, in part, to the level of arousal or engagement produced by the task. Hebb (1955) originally introduced the concept of the influence of arousal by defining task performance as a normally distributed bell-shaped curve with respect to arousal. As a result, this theory produced a value for arousal at which performance is optimal, referred to as the "optimal level of arousal." Correspondingly, the theory posits that tasks with extremely low or high levels of engagement will result in degraded performance. Because both the target acquisition and

PVT tasks were relatively simplistic and repetitive, it is theorized that the level of arousal was below the optimal level. Due to the relative complexity and higher difficulty of the UAV landing task, it is highly plausible that the task engendered an elevated level of engagement/arousal (closer to the “optimal level of arousal”) thereby benefiting performance. This increased performance may have masked any negative consequences of sleep deprivation. However, it should be noted that such effects are often short-lived.

Another potential factor resides in the fact that the subjects utilized in this experiment were not pilots and did not have any previous aircraft piloting experience. Although each subject was trained to the point that his/her performance reached a plateau, several subjects never reached ideal proficiency. In other words, their performance remained consistently poor during training. Because the training criteria was to reach the individual’s performance plateau and not to reach a specific level of proficiency, the subjects were not dismissed from participation in the study. During data collection, these subjects exhibited large variations in performance that may have also contributed to a masking of fatigue consequences.

Significant increases in the proportion of eye closure time metric were found via t-tests hours prior to significant changes in task performance for both the PVT and target acquisition tasks. Therefore, the results support the notion that eye metrics can be utilized to predict the onset of fatigue *before* the negative consequences begin to manifest. Combined with the fact that the cameras are located near the eye (as opposed to on the panel or dash) thereby reducing the likelihood of losing track of the pupil/eye closure, the results suggest that the technology has excellent potential in providing fatigue awareness and prediction capability in Air Force environments. Nevertheless, it should be noted that the current EC6 technology was designed for laboratory use only and would need to be integrated into the pilot’s life support equipment (e.g. helmet or oxygen mask) to be useful in the operational environment. Likewise, custom algorithms designed to monitor eye closure duration in near real-time would need to be designed and implemented.

The ApEn analysis illustrated that there is a statistically significant reduction in the relative disorder of the pupil tracking signal as the level of sleep deprivation increases. This supported the hypothesis that as increased levels of fatigue are coupled with reductions in complexity or irregularity of the eye movements partially as a result of an increased proportion of time staring at the screen rather than rapidly searching for targets. Lower complexity of the tracking signal translates into lower values of ApEn. Also the use of the ApEn real time measure may be a predictor of an increased fatigue state. Thus a negative rate of change of ApEn may indicate the onset of fatigue. The time rate of change of ApEn has also been documented to be an excellent predictor in other settings such as G-induced loss of consciousness (Repperger, Alberty, and Tripp, 2004; Repperger, Alberty, Tripp, and McKinley, 2004) and as a leading indicator of a possible change in financial stock market data (Pincus and Kalman, 2004). It is believed that the ApEn metric will be particularly useful in the flight environment due to the fact that pilots must constantly perform instrument crosschecks, search for targets, etc. which require continuous eye movements. The results of this experiment indicate that these movements will decline in complexity as fatigue begins to set in. This is similar to findings by Russo et al. (1999) that indicated decreases in saccadic velocity are correlated with increased sleep deprivation.

CONCLUSIONS AND FUTURE RESEARCH

The results of the experiment provide ample evidence that the eye metrics, such as the proportionate total time of eye closure and approximate entropy, can be used to indicate the onset of fatigue in advance of significant changes in operator performance.

The present results using ApEn indicate this metric is a strong leading indicator of a change in state (it has a high sensitivity). Having a high sensitivity, however, may lead to many false positives and the use of ApEn may be hampered by having a low specificity. Lack of specificity in this case would mean the ApEn indicator may change but the human may not be in a fatigue state. The prior work on ApEn has shown a high sensitivity of ApEn as a leading indicator, but additional studies should also examine the specificity aspects of using ApEn. However, it should be noted that the EyeCom, Inc., technology and software compliments similar attempts to monitor operator alertness such as those found in Ji, Zhiwei, & Lan (2004). In fact, it is suggested that the metrics presented in this study should be employed in concert with additional metrics, such as saccadic velocity described by Russo et al. (2003) and increases in pupil constriction described by Russo et al. (1999), in a final system development. Such additions would serve to reduce the risk of false positive errors.

Finally, because the cameras were mounted on eyeglass-like frames, the system was able to continuously monitor the eye throughout all sessions. Overall, the system consistently and reliably monitored the subject's eye, thereby eliminating field of view constraints characteristic of dash-mounted systems. It should be noted that the analyses used in the described experiments were performed post hoc and merely provide evidence that metrics exist that can be monitored in real time or near real time to predict negative performance effects from stressors such as fatigue. Although the EyeCom device does collect eye state data in real time, additional algorithm development combined with systems integration is necessary to ensure the system is usable in the flight/combat environment. The technology remains an experimental system and maintains some integration and comfort issues. However, the current design was not intended for operational use and will need to be customized for specific applications.

REFERENCES

- Caldwell, J.A. (2005). Fatigue in aviation. *Journal of Travel Medicine and Infectious Disease*, 3(2): 85-96.
- Caldwell, J.A., Caldwell, J.L., Schmidt, R.M. (2008). Alertness management strategies for operational contexts. *Sleep Medicine Review*, 12:257-273
- Dawson, D., & Reid, K. (1997). Fatigue, alcohol and performance impairment. *Nature*, 388:235.
- Dinges, D.F. (1990). The nature of subtle fatigue effects in long-haul crews. *Proceedings of the Flight Safety Foundation of the 43rd International Air Safety Seminar, Italy*. 7, Arlington, VA: Flight Safety Foundation.
- Dinges, D.F., Mallis, M.M., Maislin, G., & Powell, J.V. (1998). Evaluation of techniques for ocular measurement as an index of fatigue and as the basis for alertness management (Rep No. FHWA-MCRT-98-006).
- Dinges, D.F., & Grace, R. (1998). PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance. *Federal Highway Administration, Office of Motor Carriers* (Rep. No. FHWA-MCRT-98-006).
- Dinges, D.F., Pack, F., Williams, K., Gillen, K.A., Powell, J.W., Ott, G.E., Aptowicz, C., & Pack, A.I. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep*, 20: 267-277.
- Fletcher, A., & Dawson, D. (2001). A quantitative model of work-related fatigue: empirical evaluations. *Ergonomics*, 44(5): 475-488.
- Goode, J.H. (2003). Are pilots at risk of accidents due to fatigue?. *Journal of Safety Research*, 34: 309-313.
- Grace, R. (2001). Drowsy driver monitor and warning system. *Proceedings of Driving Assessment 2001: International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.
- Hebb, D.O. (1955). „Drives and the C.N.S. (Conceptual Nervous System)“. *Psychological Review*, 62: 243-254.
- Institute of Medicine (2004). Metabolic monitoring technologies for military field applications. Washington, DC: Institute of Medicine.
- Ji, Q., Zhiwei, Z., Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE Transactions on Vehicular Technology*, 53(4): 1052-1068.
- Mallis, M.M., Maislin, G., Powell, J.W., Konowal, N.M., & Dinges, D.F. (1999). Perclos predicts both PVT lapse frequency and cumulative lapse duration. *Sleep*, 22(1): 149.
- Mallis, M.M., Neri, D.F., Colletti, L.M., Oyung, R.L., Reduta, D.D., Van Dongen, H., & Dinges, D.F. (2004). Feasibility of an automated drowsiness monitoring device on the flight deck. *Sleep*, (Suppl. 27): A167.
- McFarland, R.A. & Edwards, H.T. (1953). Human factors in air transportation. *Occupational Health and Safety*, New York; McGraw-Hill.

- Perry, I.C. (Ed.). (1974). Helicopter aircrew fatigue. *AGARD (Advisory Rep. No. 69). Neuilly sur Seine, France: Advisory Group for Aerospace Research and Development.*
- Pincus, S. & Kalman, R.E. (2004). Irregularity, volatility, risk, and financial market time series. *Proceedings of the National Academy of Sciences, 101(38): 13709-13714.*
- Repperger, D.W., Frazier, J.W., Popper, S., & Goodyear, C. (1990). Attention anomalies as measured by time estimation under G stress. *Biodynamics and Bioengineering Division, Wright-Patterson Air Force Base, Ohio.*
- Repperger, D.W., Albery, W.B. & Tripp, L.D. (2004). Approximate entropy as an assessment tool for system complexity and performance valuation in human-machine systems. *Proceedings of the 9th IFAC Symposium on Human-Machine Systems, Sept 7-9, 2004, Georgia tech., Atlanta, Georgia.*
- Repperger, D.W., Albery, W.B., Tripp, L.D., & McKinley, R.A. (2004). Using real time measures (approximate entropy) to estimate the cognitive state of the pilot. *Proceedings of the 29th Annual Dayton-Cincinnati Aerospace Science Symposium, March 9, 2004, Dayton, Ohio.*
- Rosekind, M.R., Graeber, R.C., Dinges, D.F., Connell, L.J., Rountree, M.S., Spinweber, C.L., & Gillen, K.A. (1994). Crew factors in flight operations IX: Effects of planned cockpit rest on crew performance and alertness in long-haul operations NASA Technical Memorandum No. 108839). *Moffet Field, CA: National Aeronautics and Space Administration, Ames Research Center.*
- Russo, M., Thomas, M., Thorne, D., Sing, H., Redmond, D., Rowland, L., Johnson, D., Hall, S., Kirchmar, J., & Balkin, T. (1999). Sleep deprivation related changes correlate with simulated motor vehicle crashes. In R. Carroll (Ed.) *Ocular measures of driver alertness. Technical conference proceedings (FHWA Technical Rep. No. FHWA-MC-99-136), Washington DC, Federal Highway Administration, Office of Motor Carrier and Highway Safety, 119-127.*
- Russo, M., Thomas, M., Thorne, D., Sing, H., Redmond, D., Rowland, L., Johnson, D., Hall, S., Kirchmar, J., & Balkin, T. (2003). Oculomotor impairment during chronic partial sleep deprivation. *Clinical Neurophysiology, 114: 723-726.*
- Schmidt, C., Collette, F. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology, 24(7): 755-789.*
- Schmidtke, H. (1976). Vigilance. In: E. Simonson & P.C. Weisner (Eds). *Psychological aspects and physiological correlates of work and fatigue. Springfield, Illinois, 193-219.*
- Stern, A. (1999). Ocular based measures of driver alertness. *Ocular Measures of Driver Alertness: Technical Conference Proceedings, USA, 4-9.*
- Wierwille, W. (1999). Historical perspective on slow eyelid closure: Whence PERCLOS?. *Ocular Measures of Driver Alertness: Technical Conference Proceedings, USA, 31-51.*
- Zakay, D., Fallach, E. (1984). Immediate and remote time estimation – a comparison. *Acta Psychologica, 57: 69-81.*

BIOGRAPHIES

R. Andy McKinley is a Biomedical Engineer at the Air Force Research Laboratory's Vulnerability Analysis Branch located at Wright-Patterson AFB, OH. He received his Ph.D. in Engineering from Wright State University in 2009, and currently serves as the Human Effectiveness Team Lead for the AFRL Rotary-wing Brownout Solution Program. Additionally, Dr. McKinley is exploring non-invasive transcranial stimulation techniques to improve human cognitive performance above normal baseline values. His recent work has also focused on modeling the effects of acceleration stress on physiologic and cognitive performance. Dr. McKinley's other research interests include multisensory displays for unmanned aerial systems, assessing the limits of tactile displays, and utilizing eye metrics for human performance monitoring.

Lindsey K. McIntire is a research associate at the Air Force Research Laboratory's Vulnerability Analysis Branch located at Wright-Patterson AFB, OH. She received her B.A. in Political Science from the Wright State University in 2004. Her research interests include operator selection for unmanned aerial systems and non-invasive transcranial stimulation to improve human cognitive performance.